webm project

Google

Jerome Jiang
Marco Paniconi

VP9 Scalable Video Coding (SVC)

Jerome Jiang

jianj@google.com
Google LLC

Outline

- Introduction
- SVC (Scalable Video Coding) in VP9
- SVC Metrics Comparison v.s. VP8 Simulcast
- Temporal Denoiser for VP9 SVC

SVC (Scalable Video Coding) in VP9

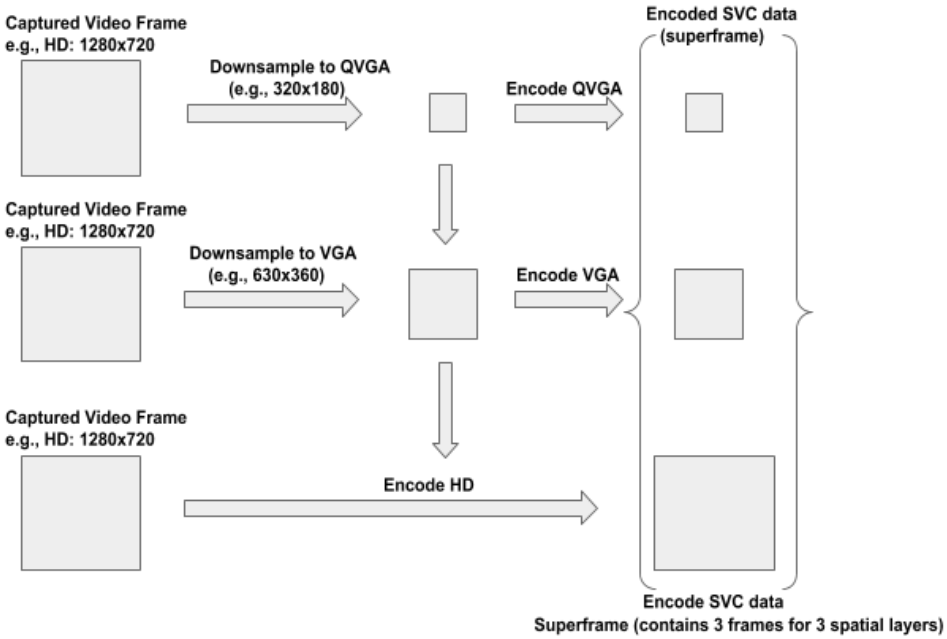
- Fully integrated in WebRTC
- Dogfood :)
- New features rolling in actively

SVC in VP9

- Unique feature of reference frame scaling
  - Spatial layers for SVC
  - Dynamic resize (change resolution within stream without key frame)
- Intra-only frame
- Multiple spatial & temporal layers
- Change layer pattern on the fly (flexible SVC mode)
- Long Term Temporal Prediction
- Noise estimation & denoising
  - All spatial layers

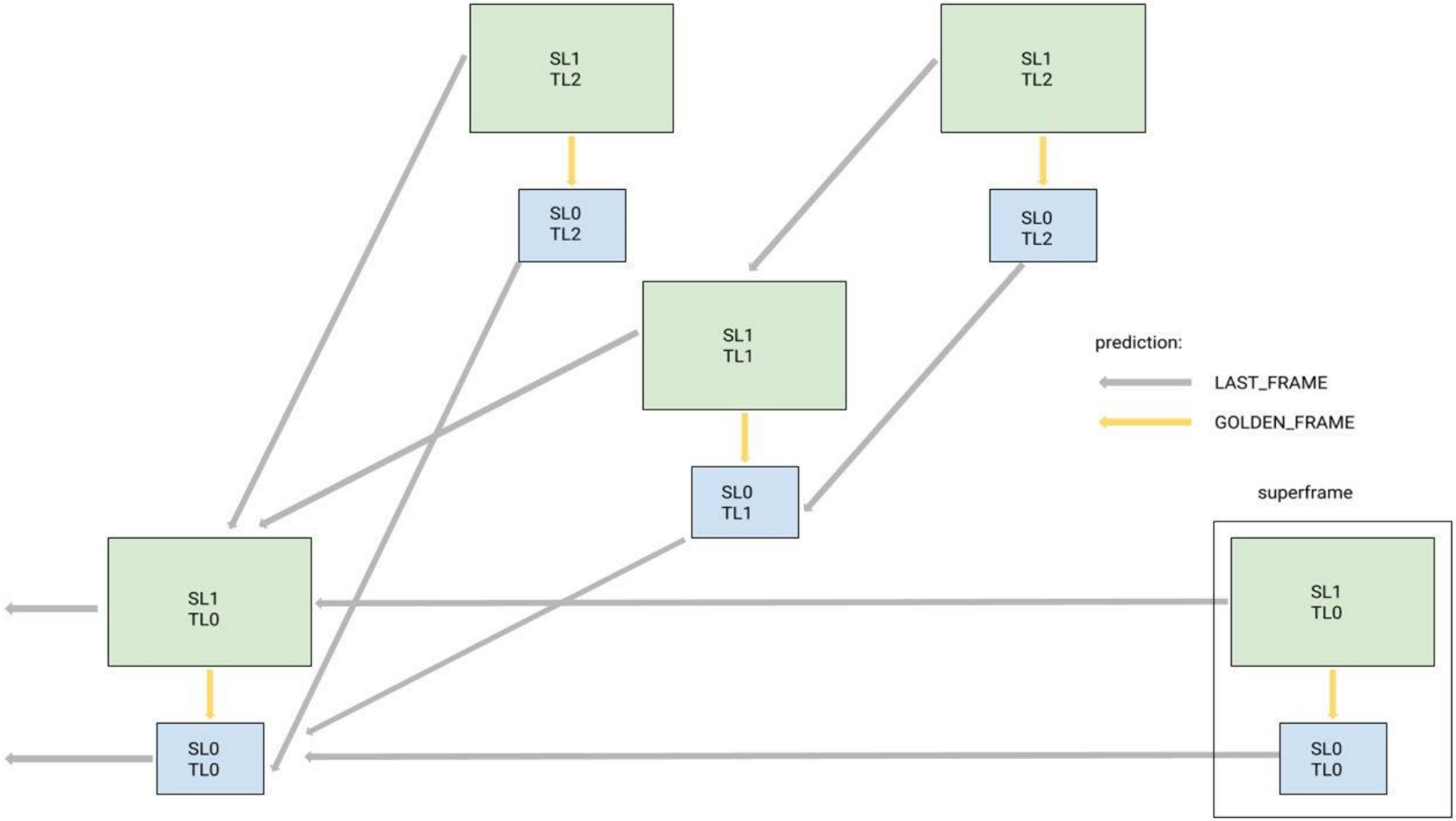
SVC Superframe

- A superframe is a frame packet containing all spatial layers.
- Downsample to lowest resolution first then encode
- Higher resolution frames predict from lower resolution ones



3 Spatial Layers

SVC Patterns - 2 Spatial Layers, 3 Temporal Layers



SVC Reference frame buffer and refresh

- ALTREF reference frame buffer is used in SVC.
- 2SL 3TL example:

	SL0 TL0		SL1 TL0		SL0 TL2		SL1 TL2		SL0 TL1		SL1 TL1		SL0 TL2		SL1 TL2	
	B	R	B	R	B	R	B	R	B	R	B	R	B	R	B	R
0	L	✔	G		L				L							
1			L	✔	G		L		G		L		G			
2					A	✔	G		A	✔	G		L	✔	G	
3							A				A	✔			L	

B = Buffer index.
R = Refresh.
L = LAST_FRAME.
G = GOLDEN_FRAME.
A = ALTREF_FRAME.

SVC Interlayer Prediction

- Users have control about inter-layer prediction (configurable)
- Several modes
  - `INTER_LAYER_PRED_ON`
    - Default mode, interlayer prediction always on.
  - `INTER_LAYER_PRED_OFF`
    - Interlayer prediction always off
  - `INTER_LAYER_PRED_OFF_NONKEY`
    - Interlayer prediction off for non keyframes (K-SVC)
  - `INTER_LAYER_PRED_ON_CONSTRAINED`
    - Inter-layer prediction is on on all frames, but constrained such that any layer S (> 0) can only predict from previous spatial layer S-1, from the same superframe.

SVC Interlayer Prediction

```
// Definition
typedef enum {
  INTER_LAYER_PRED_ON,
  INTER_LAYER_PRED_OFF,
  INTER_LAYER_PRED_OFF_NONKEY,
  INTER_LAYER_PRED_ON_CONSTRAINED
} INTER_LAYER_PRED;
// Set API
vpx_codec_ctx_t codec;
vpx_codec_control(&codec, VP9E_SET_SVC_INTER_LAYER_PRED,
INTER_LAYER_PRED_ON);
```

SVC Frame Dropping

Several frame dropping modes:

- `CONSTRAINED_LAYER_DROP`
  - Upper layers are constrained to drop if current layer drops.
- `LAYER_DROP`
  - Any spatial layer can drop.
- `CONSTRAINED_DROPBASE_ENCODESKIP`
  - Base spatial layer can drop, and this forces drop of all spatial layers. Enhancement spatial layer encodes a skip frame instead of dropping.

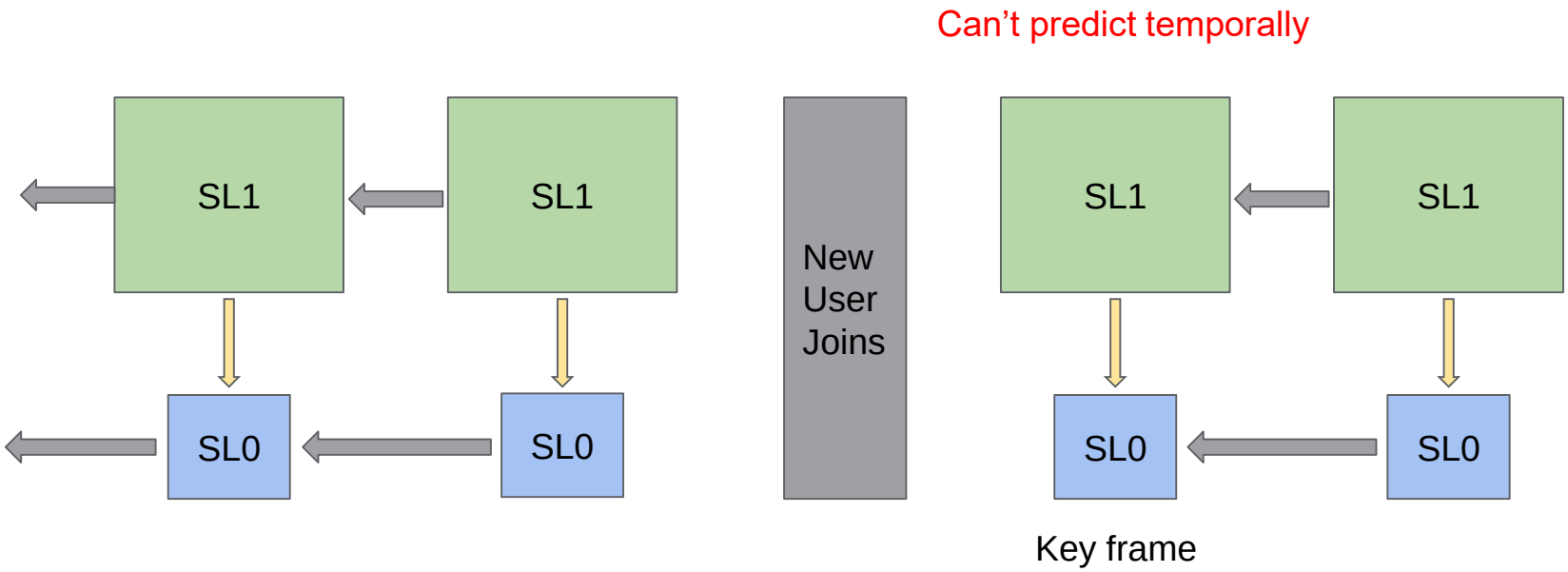
SVC Framedropping API

```
typedef struct vpx_svc_frame_drop {
  int framedrop_thresh[VPX_SS_MAX_LAYERS];
  SVC_LAYER_DROP_MODE framedrop_mode;
  int max_consec_drop;
} vpx_svc_frame_drop_t;

vpx_svc_frame_drop_t svc_drop_frame;
vpx_codec_control(&codec, VP9E_SET_SVC_FRAME_DROP_LAYER,
&svc_drop_frame);
```

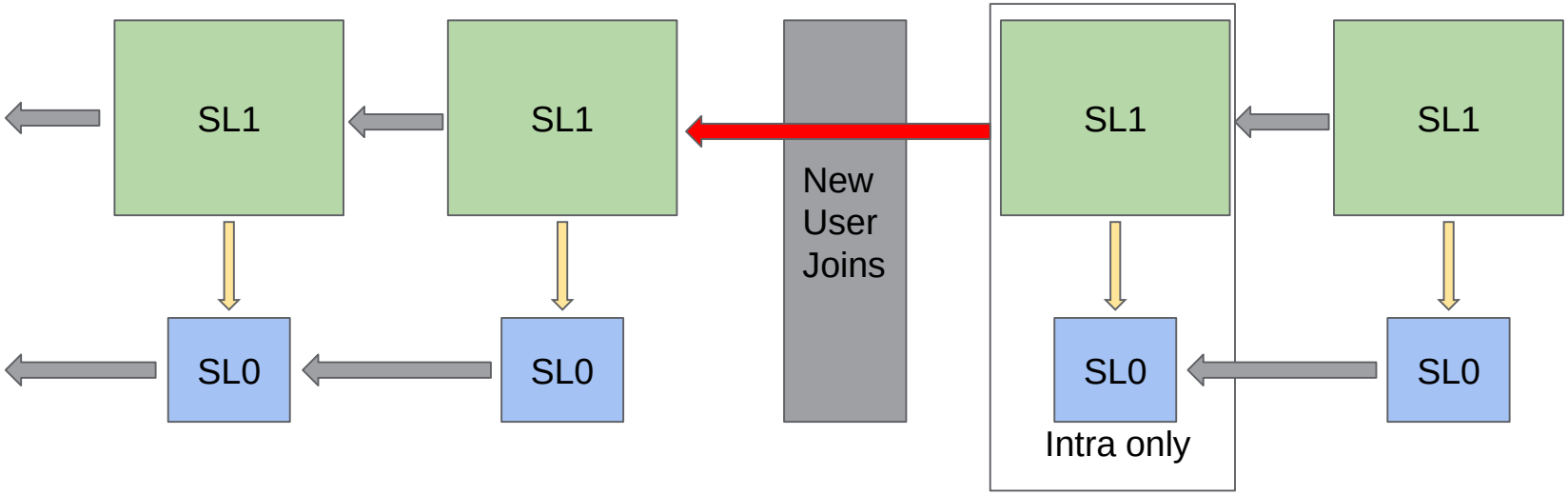
Intra-only Frame

- New user joins the group chat
  - Insert base layer as a key frame
  - All receivers need to restart the videostream



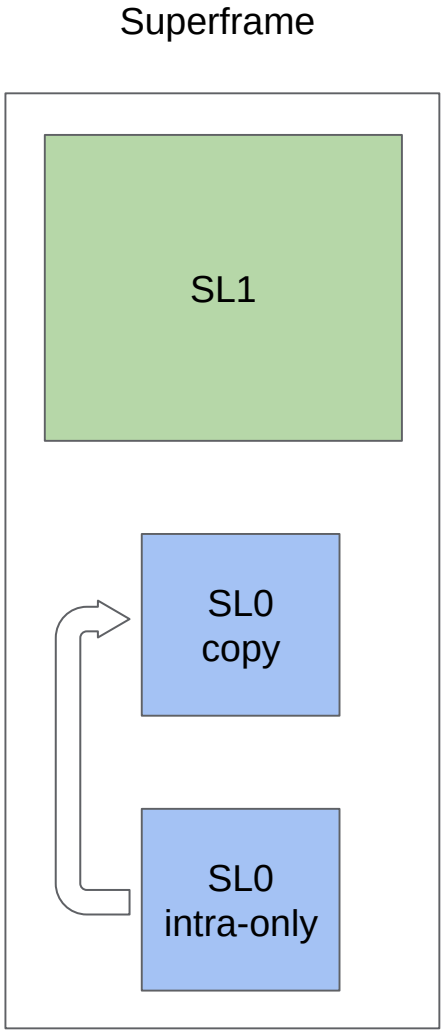
Intra-only Frame

- With intra-only frame
  - Frame encoded with intra only
  - But doesn't refresh all reference buffers
  - Must be a no show frame



Intra-only Frame

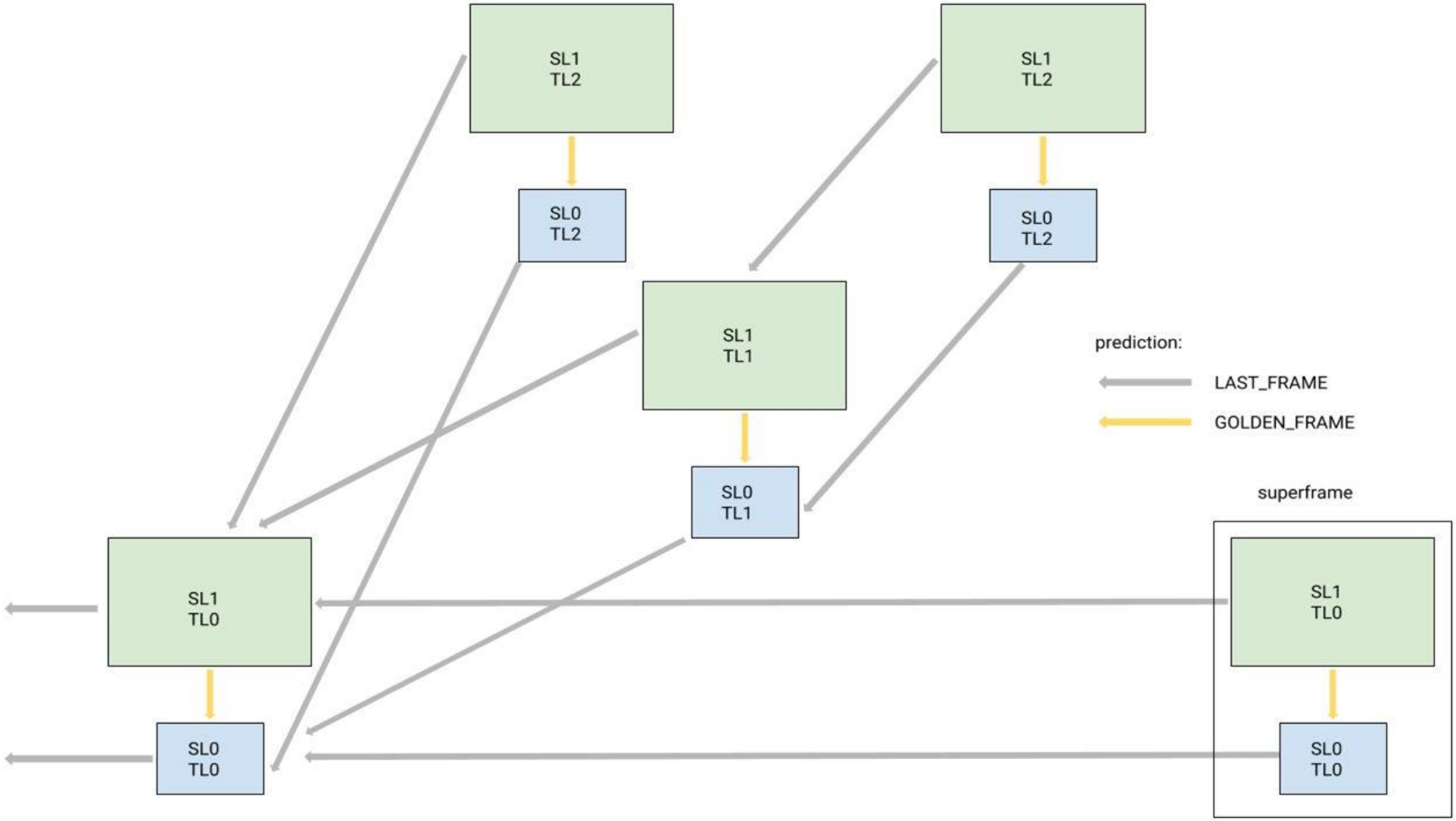
- For receivers who decode top layer
  - Can still predict temporally
  - Avoid effects of key frame
- Intra-only frame is still packed into the superframe
  - No show - (not displayed)
  - Can use flag `show_existing_frame` to copy header of intra-only frame in the superframe



Long Term Temporal Prediction

- Only applied to top spatial layer
  - Speed loss for sure, this reduces the speed loss
- Only refreshed on base temporal layer
- Use frame buffer slot 7
  - 8 in total, in most cases, slot 7 is not used.
- Quality Improvement:
  - 4% average on RTC
  - 10-12% on high motion clips

SVC Patterns - 2 Spatial Layers, 3 Temporal Layers

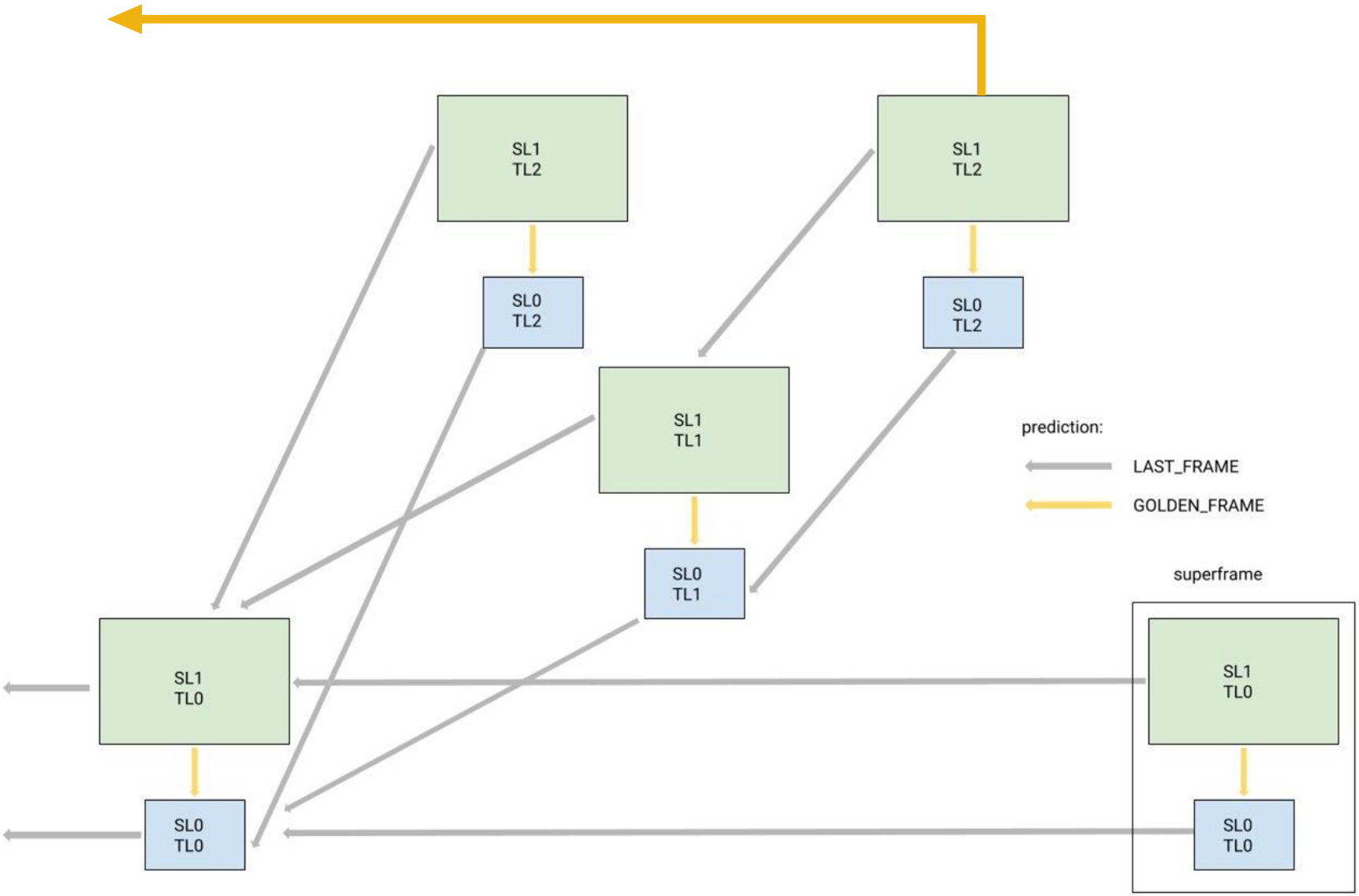


SVC Reference frame buffer and refresh

- ALTREF reference frame buffer is used in SVC.
- 2SL 3TL example:

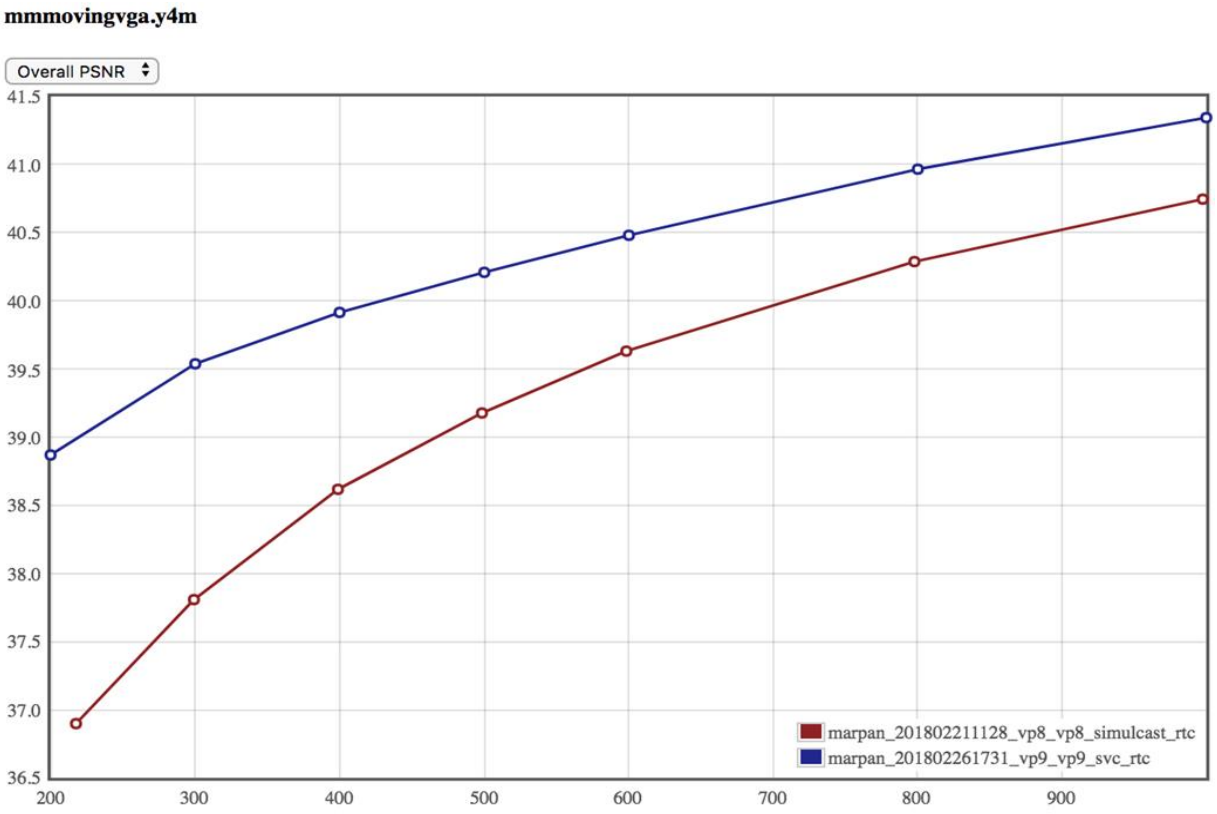
	SL0 TL0		SL1 TL0		SL0 TL2		SL1 TL2		SL0 TL1		SL1 TL1		SL0 TL2		SL1 TL2	
	B	R	B	R	B	R	B	R	B	R	B	R	B	R	B	R
0	L	✔	G		L				L							
1			L	✔	G		L		G		L		G			
2					A	✔	G		A	✔	G		L	✔	G	
3							A				A	✔			L	

B = Buffer index.
R = Refresh.
L = LAST_FRAME.
G = GOLDEN_FRAME.
A = ALTREF_FRAME.

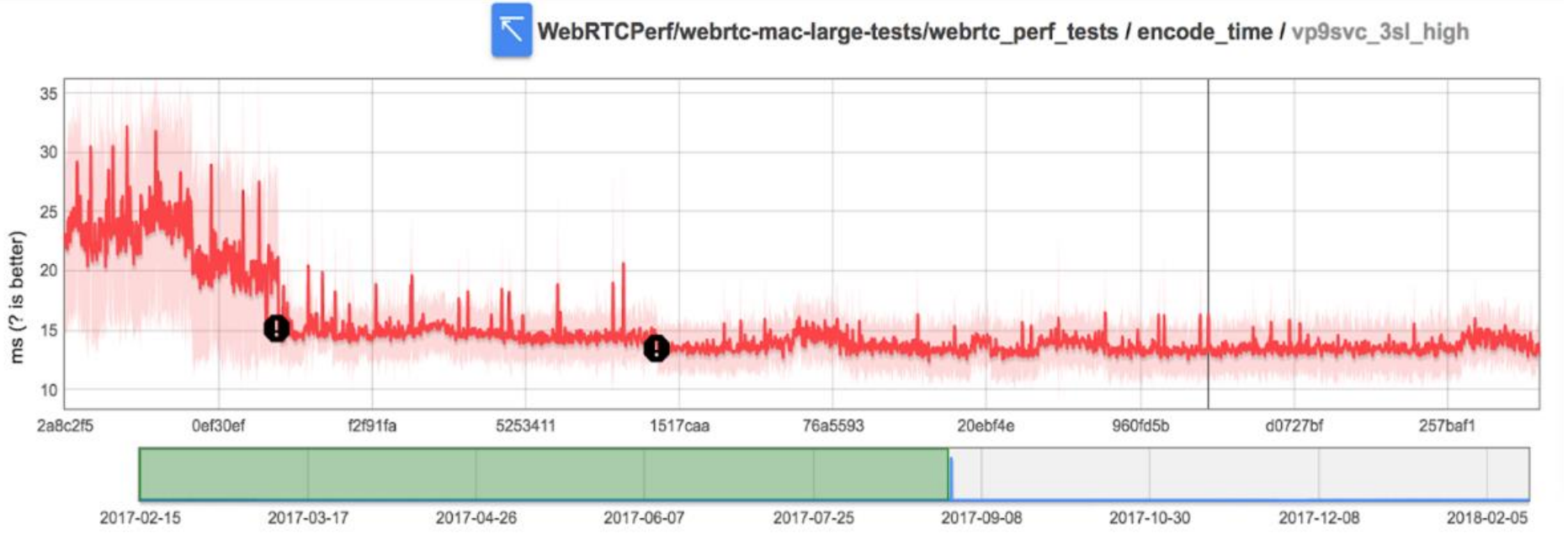

SL1
TL2
SL1
TL2
SL0
TL2
SL0
TL2
SL1
TL1
SL0
TL1
prediction:
LAST_FRAME
GOLDEN_FRAME
superframe
SL1
TL0
SL0
TL0
SL1
TL0
SL0
TL0

VP9 SVC v.s. VP8 Simulcast

File	Match	Problem	avg_psnr:
dark720p.y4m	✓	8/0	-54.809
desktop2360p.y4m	✓	7/0	-39.724
desktop360p.y4m	✓	7/0	-58.236
fourpeople720p.y4m	✓	8/0	-38.593
gipsrecmotion720p.y4m	✓	8/0	-17.756
gipsrestat720p.y4m	✓	8/0	-34.539
jimredvga_25fps.y4m	✓	7/0	-47.880
kirlandvga.y4m	✓	7/0	-60.486
marcooffice720p.y4m	✓	8/0	-28.842
mj1vc720p.y4m	✓	8/0	-40.792
mj2vc720p.y4m	✓	8/0	-29.558
mj3vc720p.y4m	✓	8/0	-35.089
mj4vc720p.y4m	✓	8/0	-25.326
mmmovingvga.y4m	✓	7/0	-42.653
mmstionaryvga.y4m	✓	7/0	-51.641
niklas720p.y4m	✓	8/0	-27.064
niklasvga.y4m	✓	7/0	-35.198
still_bright_360_640.y4m	✓	7/0	-40.011
tacomanarrowsvga.y4m	✓	7/0	-70.085
tacomascmvvga.y4m	✓	7/0	-39.452
testnoise720p.y4m	✓	8/0	-59.004
thaloundeskmtgvga.y4m	✓	7/0	-45.923
vidyo1_1280x720_60.y4m	✓	8/0	-28.229
vidyo3_1280x720_60.y4m	✓	8/0	-24.326
vidyo4_1280x720_60.y4m	✓	8/0	-38.088
{OVERALL}	✓	None	-40.532



VP9 SVC Speed up

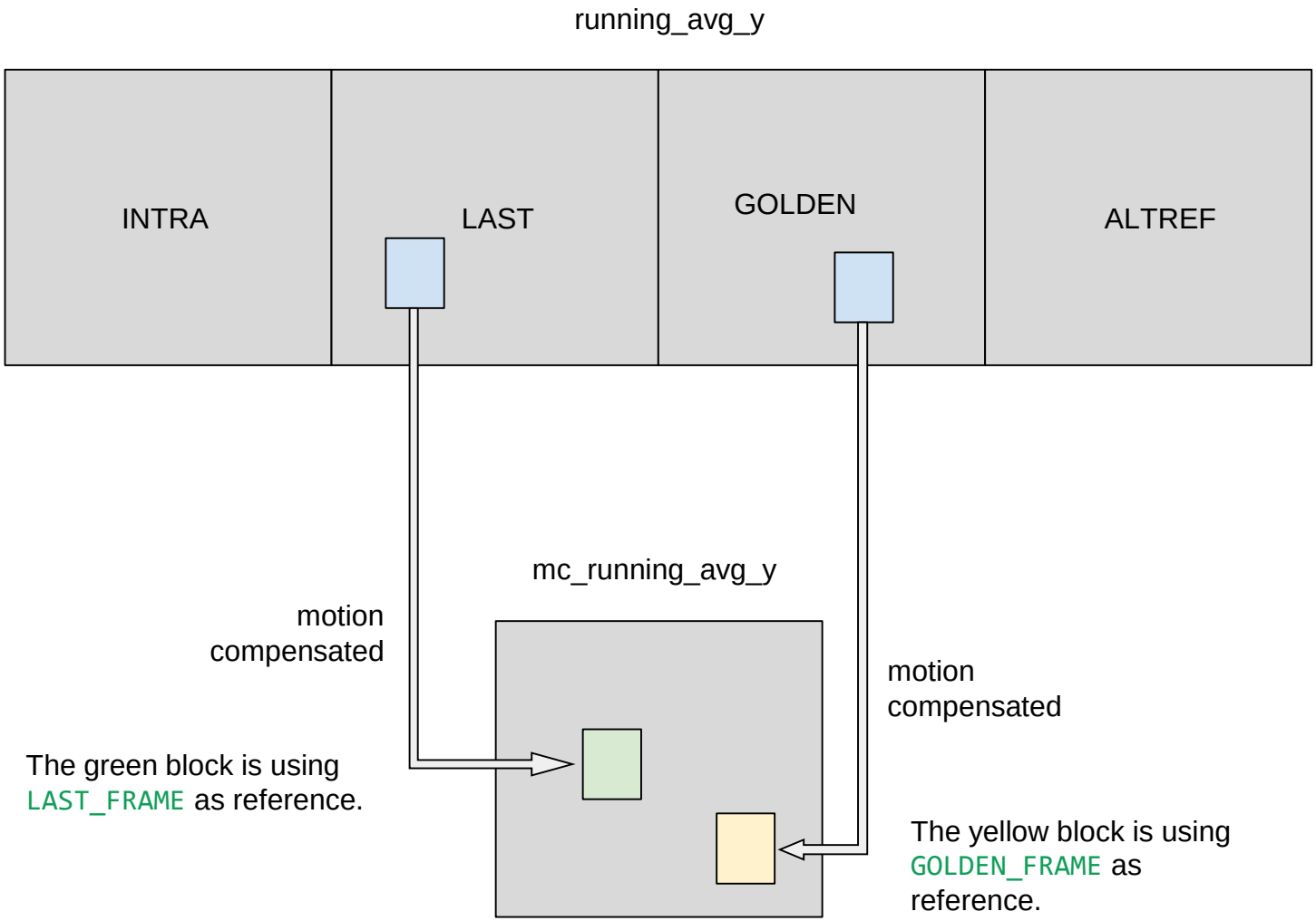


Overall 45% speed up on HD (720p).

VP9 Temporal Denoiser

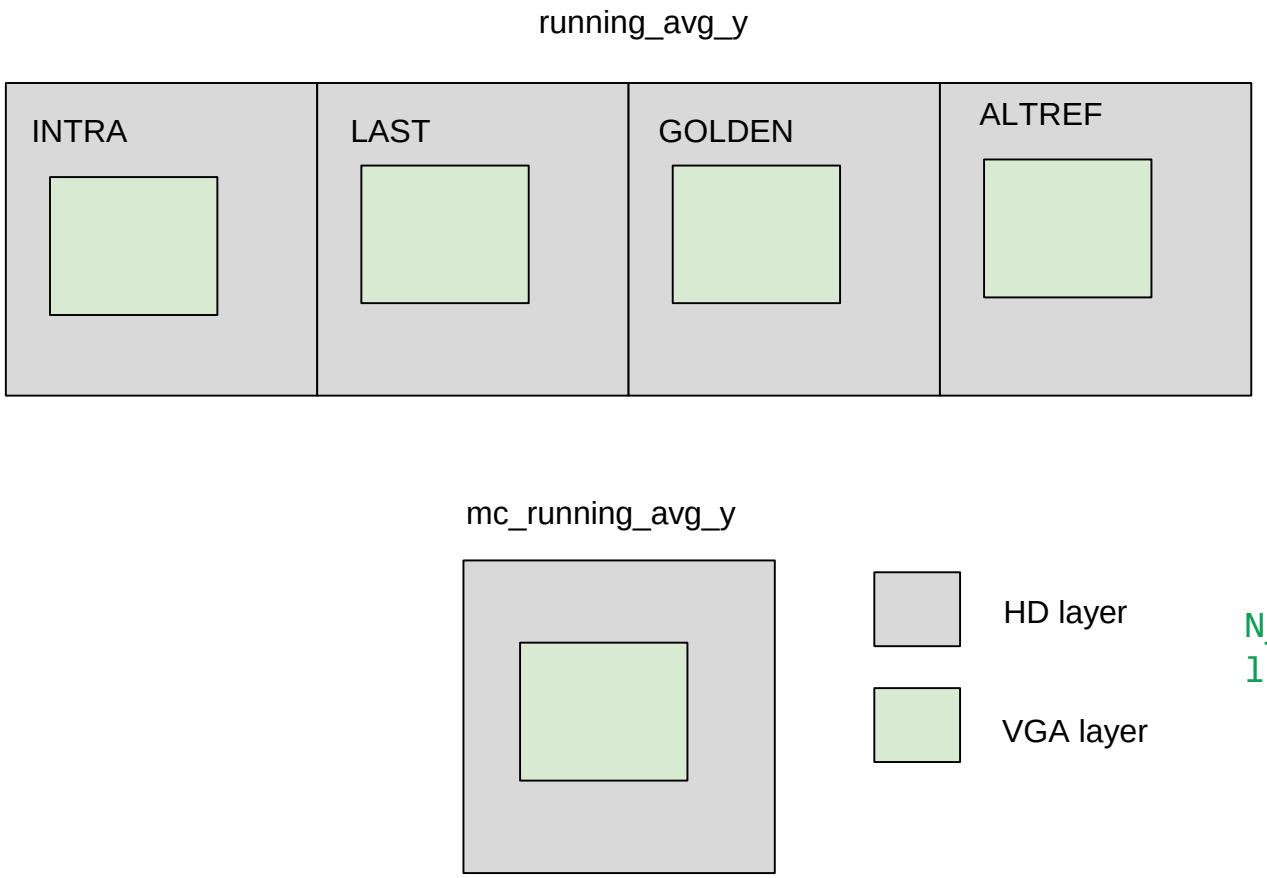
- Noise estimation
- Denoiser will decide according to noise estimation
  - If noise level is low - don’t denoise even if denoising enabled by user
  - If noise level is high
    - Perform motion compensation - return two values
      - `COPY_BLOCK` - Copy block from source without denoising
      - `FILTER_BLOCK` - Denoise the source

Denoiser Frame Buffer

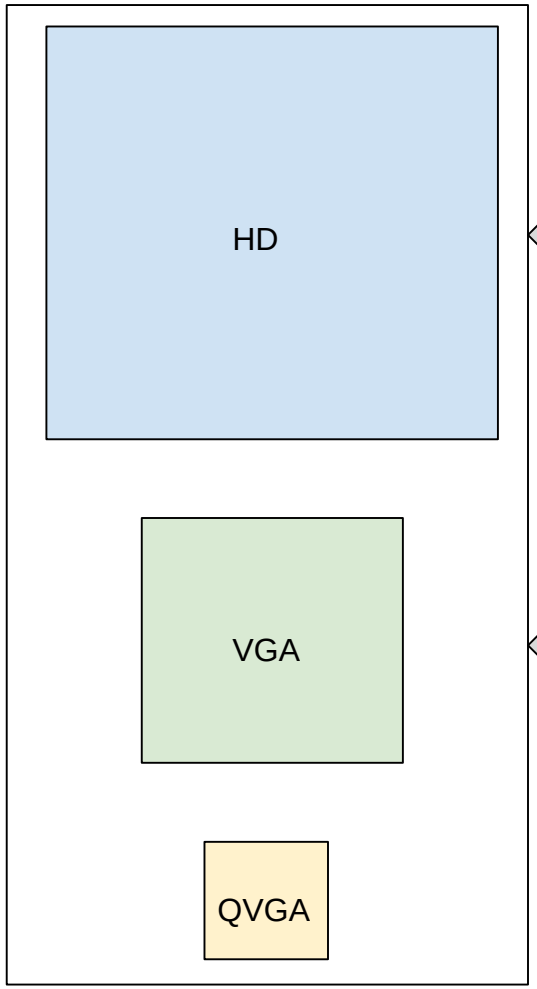


Denoiser in SVC

More frame buffers: `N_{reference frames} x N_{denoise spatial layers}`



Denoiser in SVC



first_layer_denoise

```
vpx_codec_control(&codec,
VP9E_SET_NOISE_SENSITIVITY, 1)
```

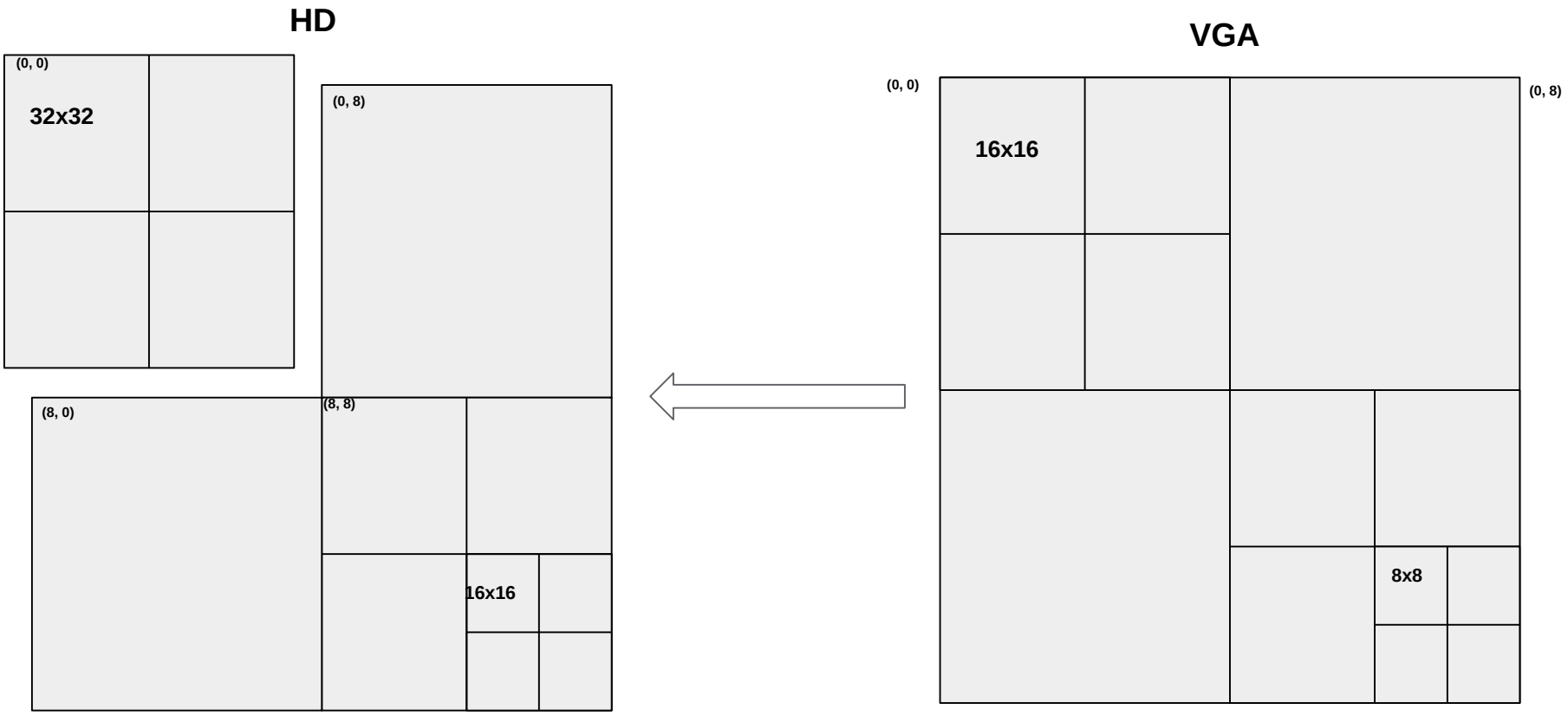
Only denoise top layer,
`N_{denoise spatial layers} = 1`

first_layer_denoise

```
vpx_codec_control(&codec,
VP9E_SET_NOISE_SENSITIVITY, 2)
```

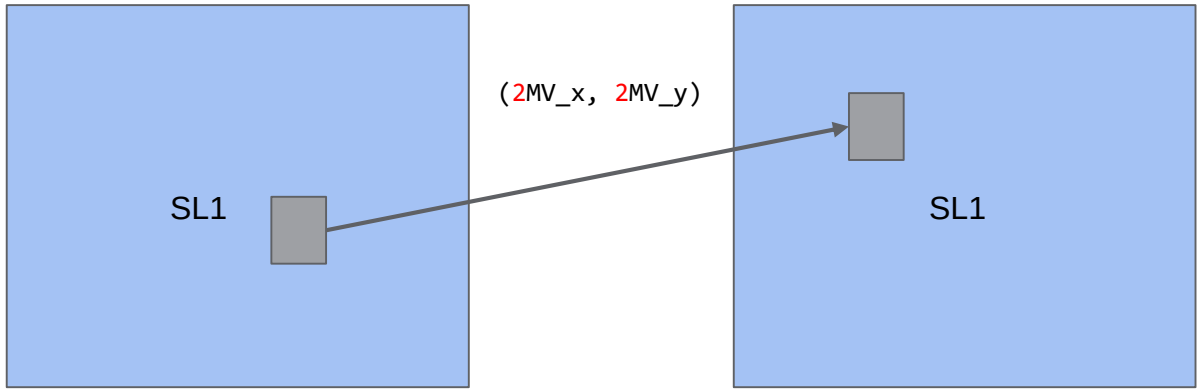
Denoise top 2 layers,
`N_{denoise spatial layers} = 2`

Scale partitioning in SVC

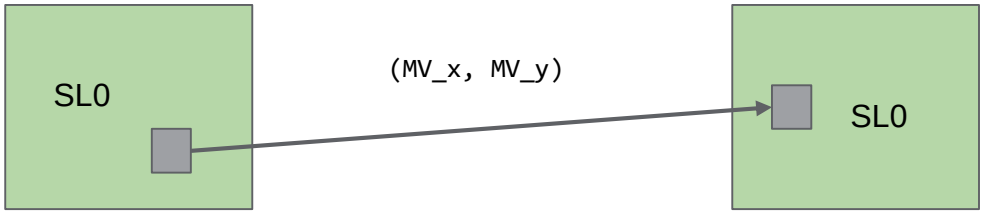


Motion Vector Reuse

Reuse motion vector from base layer for faster NEWMV search on LAST.



Smaller search area
around (2MV_x, 2MV_y)



Wrap Up

- Huge progress on VP9 SVC
  - Integrated in WebRTC for Hangouts
- VP9 SVC maintains quality advantage over VP8 simulcast
- Denoiser improves quality under noisy situations

Thank you!